ASSIGNMENT 6

RESIGRATION ID: SIRSS2101

NAME: GAURAV GANESH BHOGALE.

SUBJECT: REGRESSION

Q1. Calculate/ derive the gradients used to update the parameters in cost function optimization for simple linear regression.

The equation for simple regression is y= $a1 * x + a0$

The cost or error(e) = $y \wedge - y$

For n data points:

$$f(a) = \frac{1}{n} \sum_{i=1}^{n} (y^{\wedge} - y)^2$$

$$f(a) = \frac{1}{n} \sum_{i=1}^{n} (y^{\wedge} - (a1 * x + a0))^2$$

α= learning rate or the size of the step we take towards finding the optimal fit line

$\frac{df(a)}{da0}$ $partial\ derivative\ of\ f(a)$ with respect to $a0\ will\ give\ the\ value\ of\ parameter\ a0$

$$a0 = \frac{2}{n} \sum_{i=1}^{n} (y^{\wedge} - (a1 * x + a0))$$

$\frac{df(a)}{da1}$ $partial\ derivative\ of\ f(a)\ w.r.t\ a1\ will\ give\ the\ value\ of\ parameter\ a1$

$$a1 = \frac{2}{n} \sum_{i=1}^{n} x\,(y^{\wedge} - (a1 * x + a0))$$

New $a0 = a0 - a0 * α$

New $a1 = a1 - a1 * α$

Q2. What does the sign of gradient say about the relationship between the parameters and cost function?

- The cost function is a function of the parameters and when the sign is positive then the stepwill decrease

  New m0= m0 $-$ $[+ve\ gradient] * \alpha$

- When the sign is negative then the step will increase as seen

  below:

- New m0= m0 $-$ $[-ve\ gradient] * \alpha$

  New m0= m0 + $[\ gradient] * \alpha$

Q3. Why Mean squared error is taken as the cost function for Regression problems.

For linear regression, this MSE is nothing but the Cost Function. Mean Squared Error is the sum of the squared differences between the prediction and true value. And once we have the slope and intercept of the line which gives the least error, we can use that line to predict Y.

Q4. What is the effect of learning rate on optimization, discuss all the cases?

The learning rate hyperparameter controls the rate or speed at which the model learns. Specifically, it controls the amount of apportioned error that the weights of the model are updated with each time they are updated, such as at the end of each batch of training examples.

The cost function value will be minimized rather quickly. If we take a large learning rate then the cost function value will be minimized very quickly but will settle at a value that is not the lowest.

If we take a lower than optimal learning rate, then even after substantial iterations the cost function will not minimize sufficiently and will take longer time.